# Make sense of your logs!

Britta Weber

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.

# Who am I

- elasticsearch core developer
- joined elasticsearch the company 1.5 years ago
- interested in data analysis (but really doing a little bit of everything: support, trainings, fixes, conference talks, …)
- no ops background

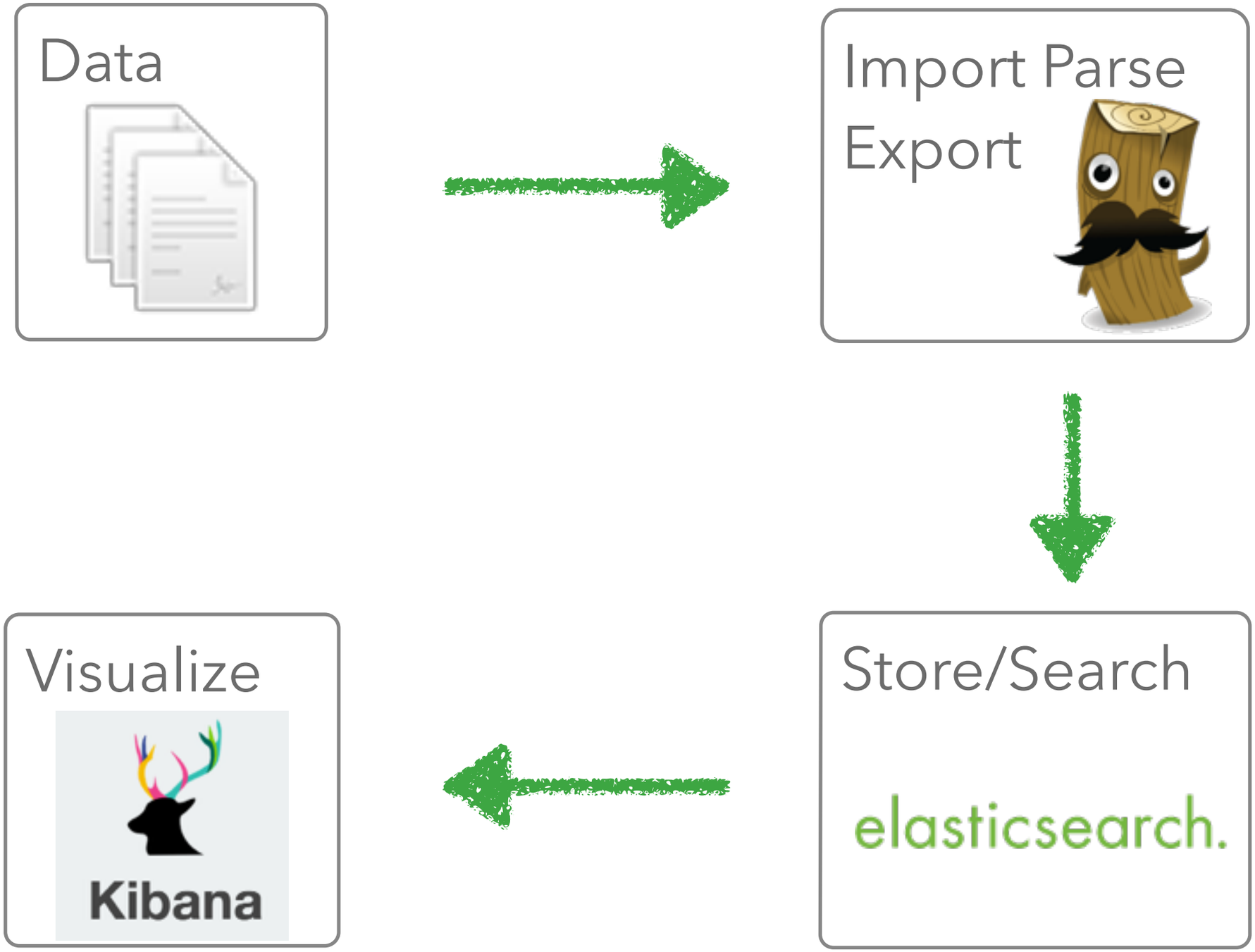
Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# The purpose of this talk

- introduction to ELK stack (elasticsearch + logstash + kibana)
- analytics with elasticsearch
- visualizing with kibana 4

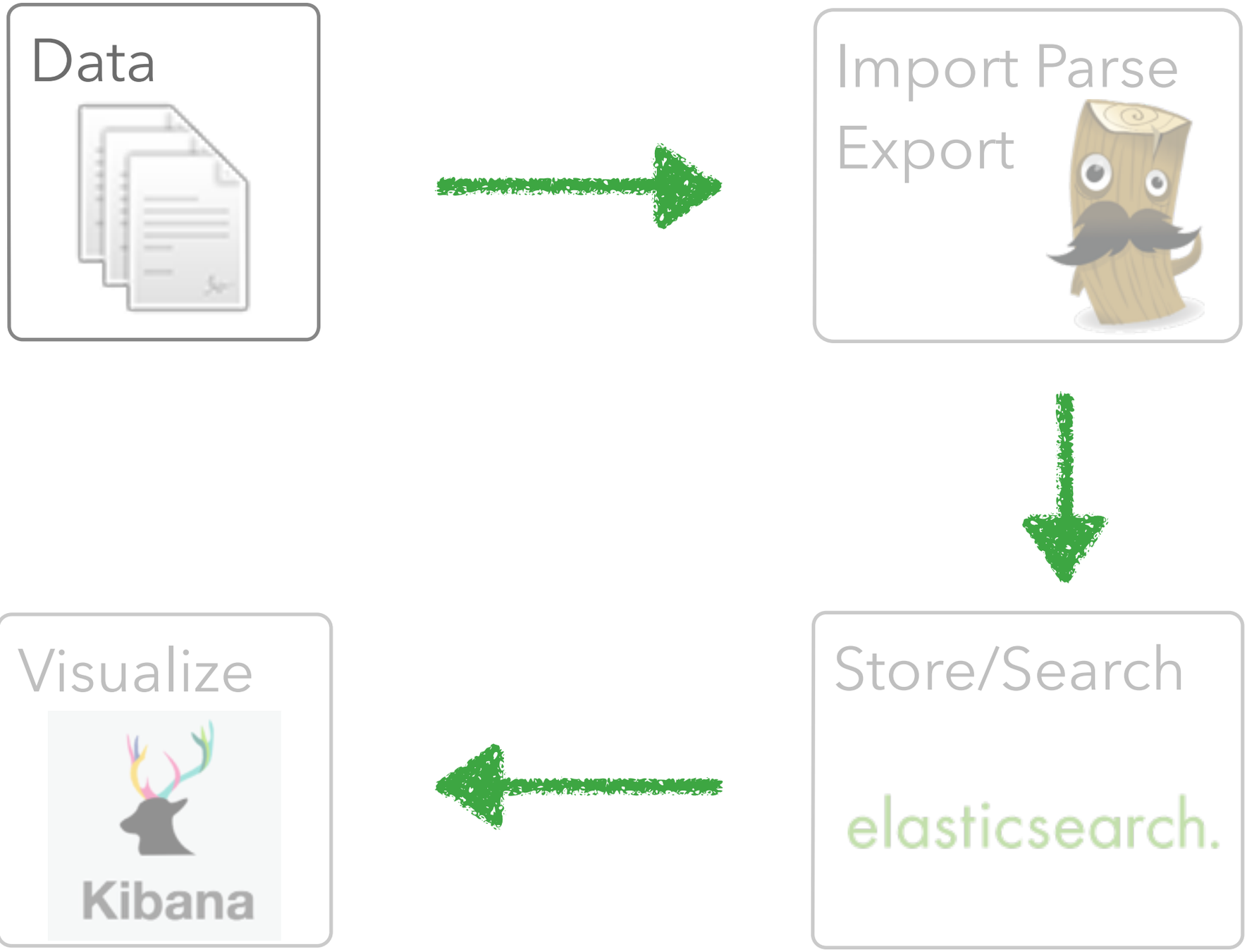
Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# The ELK stack



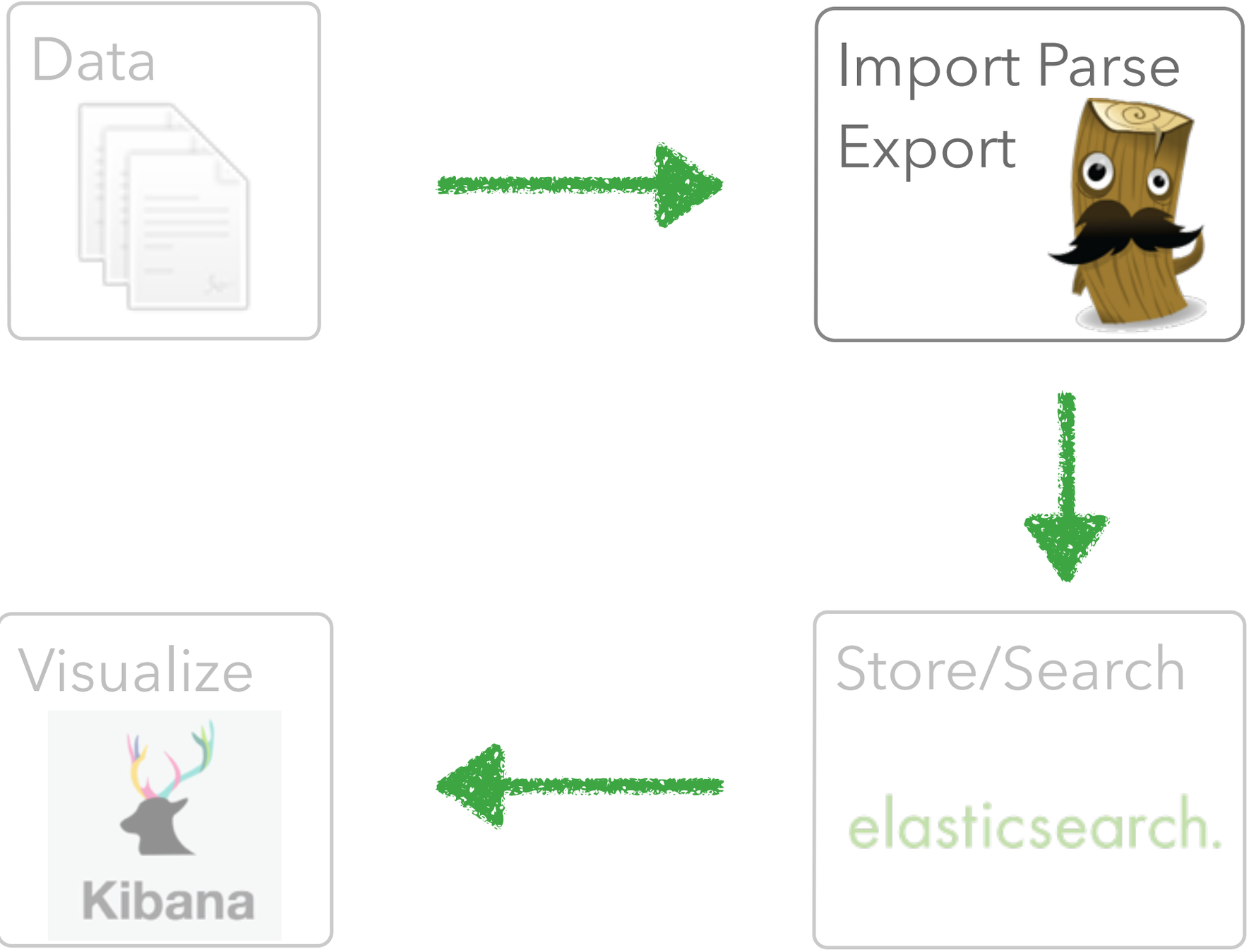

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# The ELK stack




Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# The ELK stack




Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# Logstash in 10 seconds

- Ruby app (JRuby)
- Collect, parse, enrich, store data
- Horizontally scalable
- Open source! Apache License 2.0

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# Logstash architecture

## Input

datastore

log files

files

monitoring

queues

…

## Filter

Logstash

parse, enrich, tag, drop

## Output

datastore

email

pager

chat

queues

…

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# Logstash architecture

## Input

datastore

log files

files

monitoring

queues

…

## Filter

Logstash

parse, enrich, tag, drop

## Output

datastore

email

pager

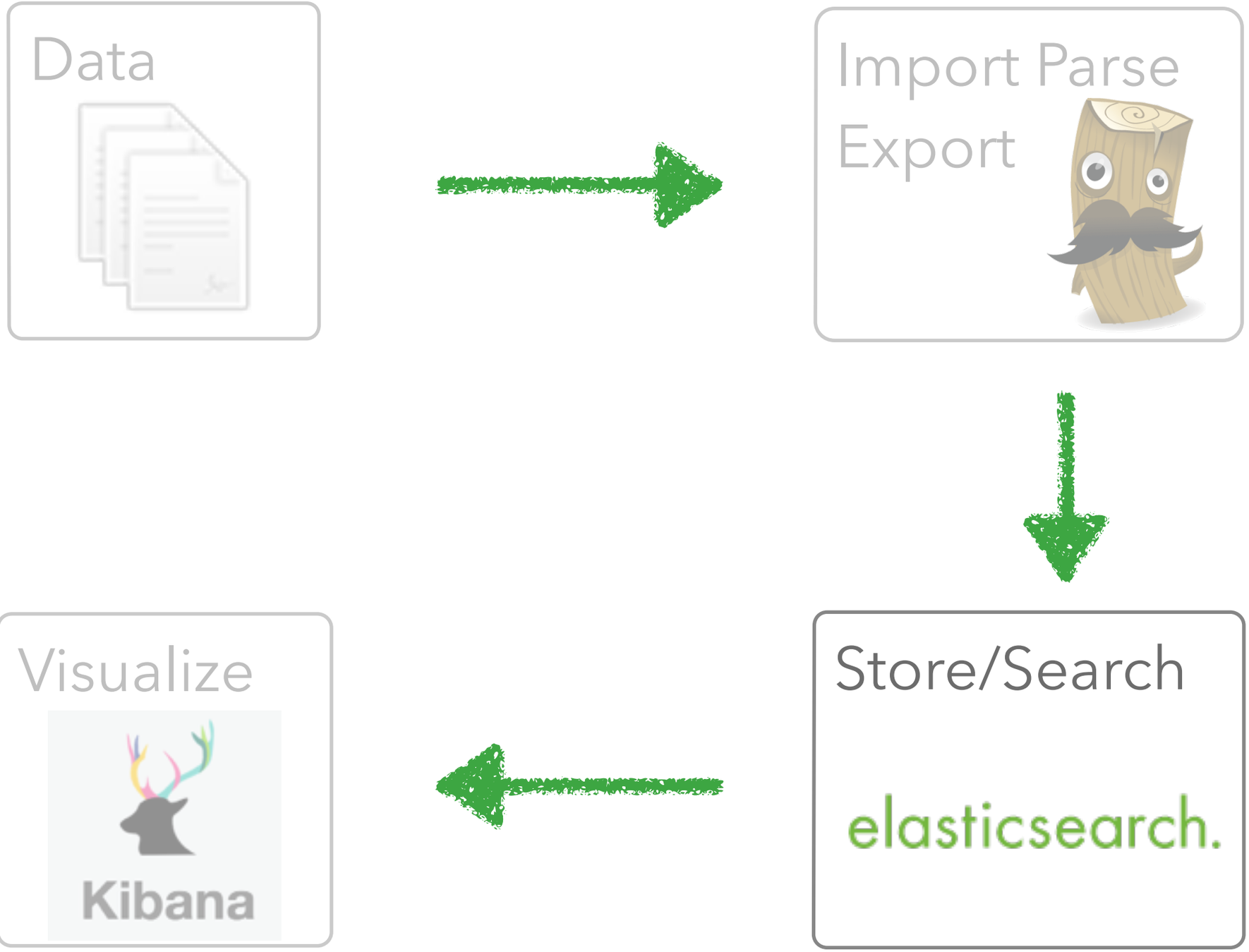
chat

queues

…

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# The ELK stack




Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# Elasticsearch in 10 seconds

- Written in Java
- Schema-free, REST & JSON based document store
- Search engine
- Distributed and horizontally scalable
- Open source! Apache License 2.0

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# “Elasticsearch the definitive guide”

Clinton Gormley & Zachary Tong

read online at www.elasticsearch.org

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.

# Index organization with logstash

- elasticsearch organizes documents in indices
- logstash per default creates indices that start with prefix logstash-DATE
- logstash organizes documents per default as time based indices - one index per day (depending on @timestamp field)

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# Analytics with elasticsearch

Summary statistics over your data

How many requests per day?

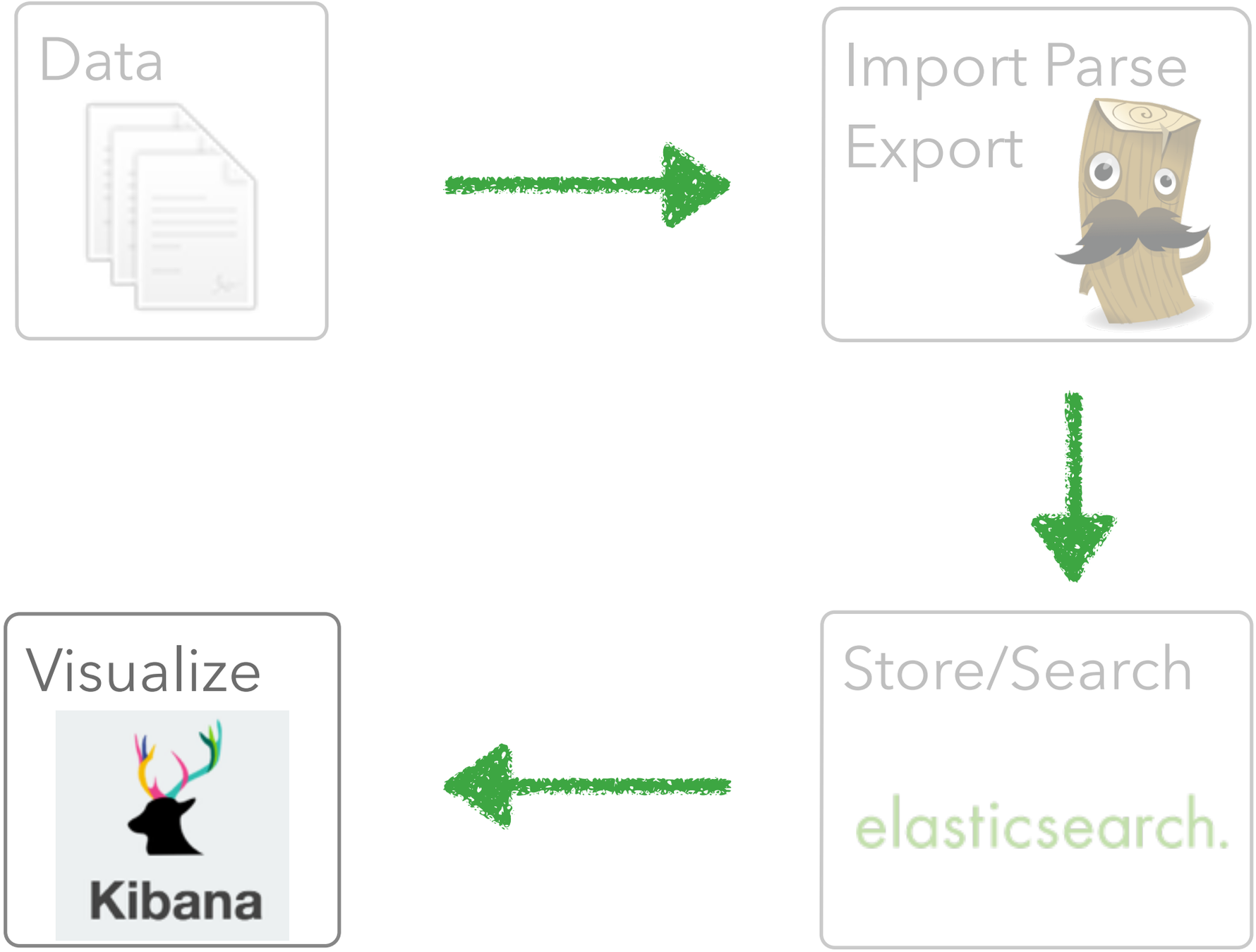
How many mozilla users visit my webpage?

Is my cpu usage increasing?

-> aggregation

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# The ELK stack




Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# Kibana in 10 seconds

- In-browser (JS, built on Angular, Flot, D3)
- Visualize data in Elasticsearch
- Open source! Apache License 2.0

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

“But wait! Did you not say Kibana 4 is a BETA?”

Yes! We are entering the…

WARNING!
DANGER
ZONE

# Data

**apache log**

http://semicomplete.com/misc/sample.log
Jordan Sissel's old homepage, hosted logstash
120k entries

**wiki access**

http://www.wikibench.eu/
VU University Amsterdam, Guillaume Pierre et al.
~ 2 month of wiki page access

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

WARNING!
DANGER
ZONE

Discover | Visualize | Dashboard | Settings | September 18th 2007, 19:53:23.433 to October 31st 2007, 07:49:50.025

*

metrics

Y-Axis — Count of documents

buckets

Split Lines — Top 20 page.raw

X-Axis — Daily @timestamp

Add Sub Aggregation

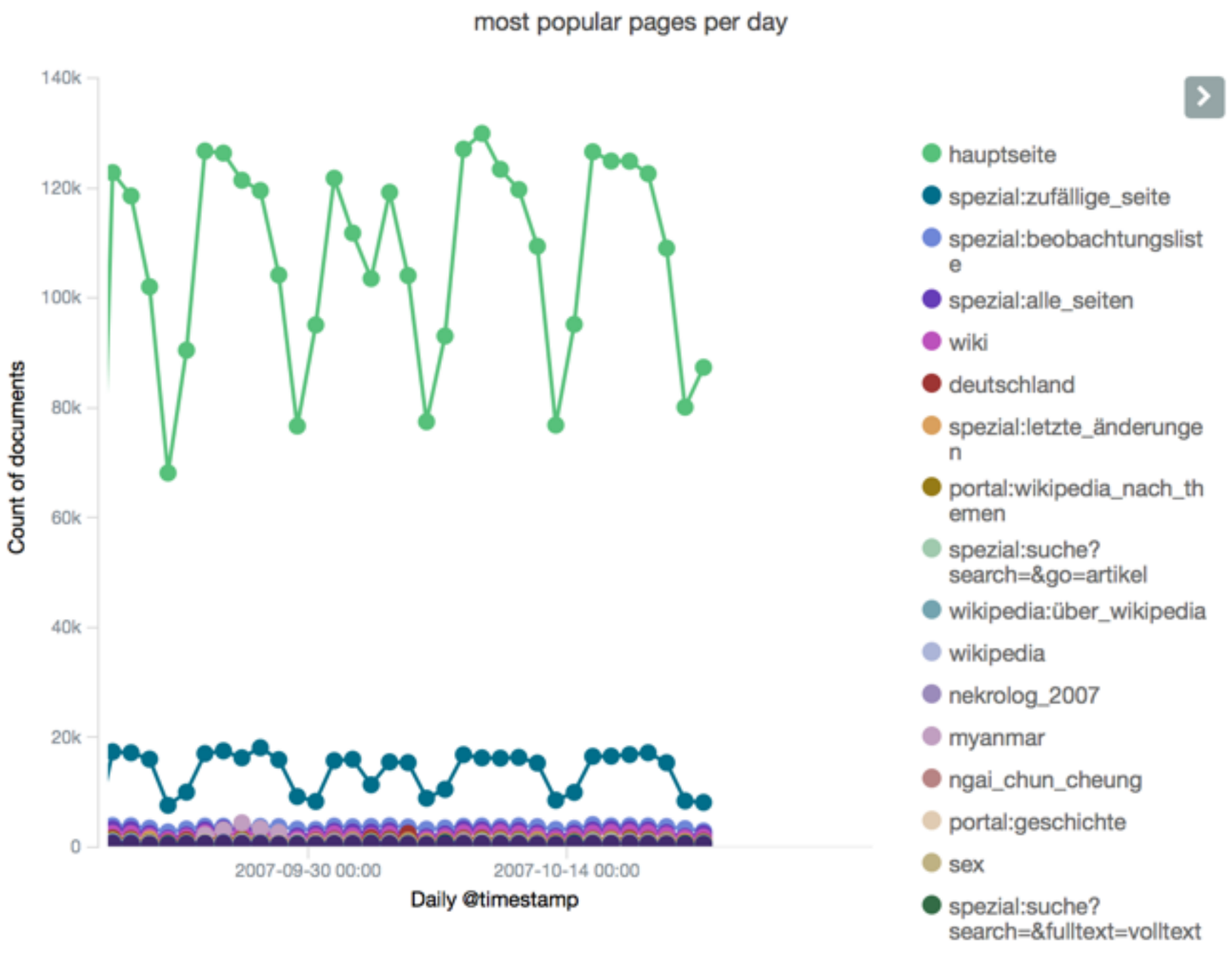
visualization options

Show Tooltip

Show Legend

Apply

Discard



Discover | Visualize | Dashboard | Settings | September 18th 2007, 19:53:23.433 to October 31st 2007, 07:49:50.025

ki") AND page.raw:(NOT "Deutschland") AND page:(NOT portal) AND page:(NOT wikipedia) AND page:(NOT special)

metrics

**Y-Axis** Count of documents

buckets

**Split Lines** Top 20 page.raw

**X-Axis** Daily @timestamp

Add Sub Aggregation

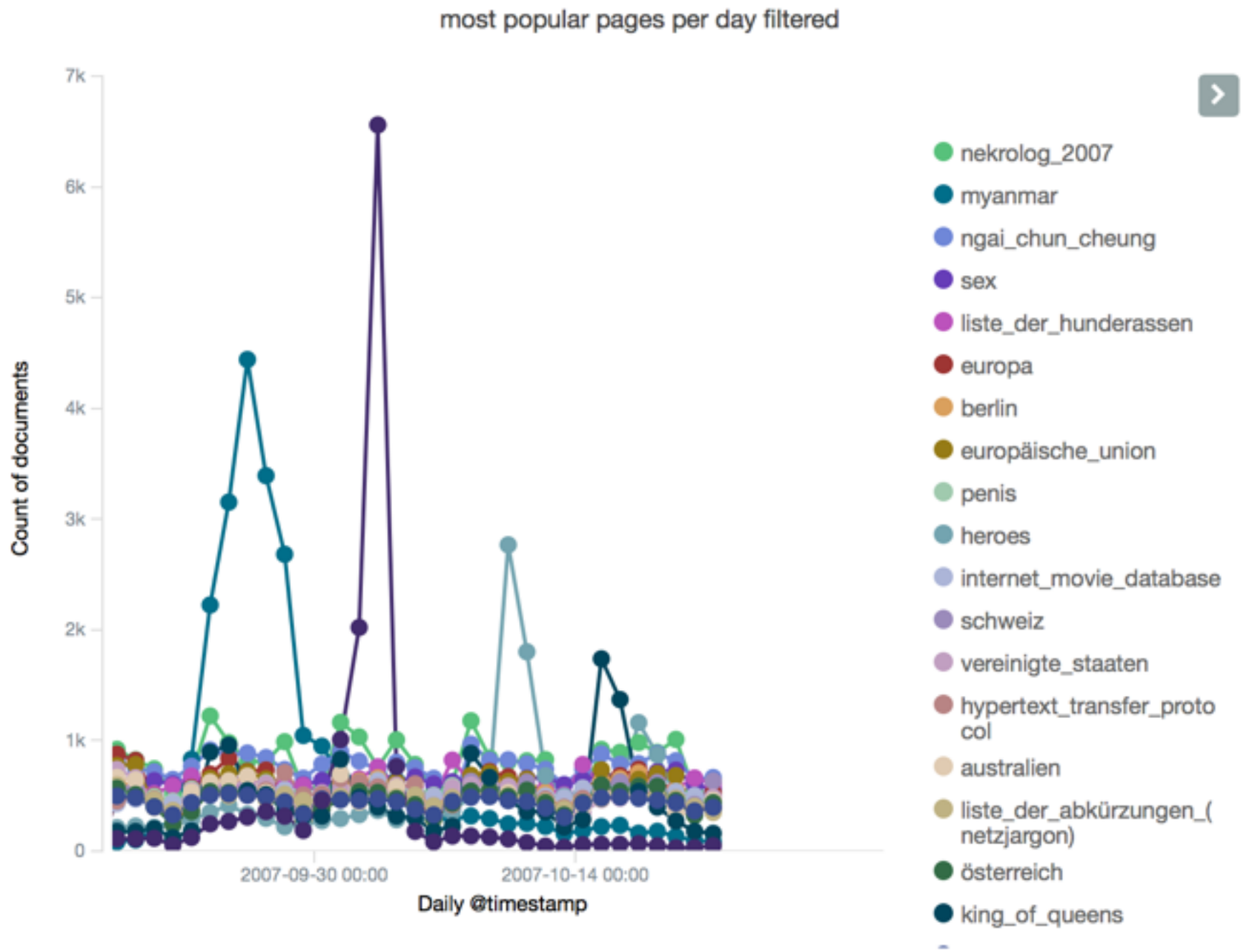
visualization options

**Show Tooltip**

**Show Legend**

Apply

Discard



Wikipedia:Fragen zur Wikipedia/Archiv/2011/Woche 05 – Wikipedia
de.wikipedia.org/wiki/Wikipedia:Fragen_zur_Wikipedia/Archiv/2011/Woche_05#Zunehmender_Missbrauch_im_Zusammenh
Reader
Apple iCloud Facebook Twitter Wikipedia Yahoo News Popular
Marvel – Sense | Wikipedia:Hauptseite/… | Wikipedia:Fragen zur… | 3sat TV-Programm Di… | Kibana 4

## Zunehmender Missbrauch im Zusammenhang mit stats.grok.se und Ngai Chun Cheung sowie Ngai.cc

In den letzten Tagen wurde hier nach den am wenigsten gelesenen Artikeln gefragt. Ich fand die Frage ganz interessant und habe versucht, mit Hilfe von Datenbankdumps und der Statistik-Rohdaten, die auch von stats.grok.se verwendet werden, eine Antwort zu finden.

Dabei sind mir die häufigen Zugriffe auf den nicht existierenden Artikel Ngai.cc aufgefallen (steht auch in den Top-1000 von Dezember 2010 auf Platz 6). Die Zugriffe finden kontinuierlich seit dem 15. September 2010 statt und haben sich seitdem von ca. 55.000 pro Tag auf heute 152.000 pro Tag gestiegert (hochgerechnet aus der Abrufstatistik von einer Stunde heute vormittag). Das waren in dieser Stunde immerhin 6338 Zugriffe (0,4% aller Seitenzugriffe auf die deutschsprachige Wikipedia in dieser Stunde).

Der Hintergrund dieses "Angriffs" erschließt sich auf ngai.cc: Hier beklagt der Betreiber die bereits lange zurückliegende Löschung seines Biografieartikels Ngai Chun Cheung (am 1.9.2006 nach eindeutiger LD mangels Relevanz gelöscht).

Bereits 2007 landete der nun gelöschte Artikel in den Top-100 auf Platz 26 (damals noch mit etwa 300 Zugriffen pro Stunde).

Da diese Aktion nun schon so lange und mit steigender Frequenz erfolgt, bin ich der Meinung, man sollte etwas dagegen tun, bevor derjenige auf die Idee kommt, sein Script nochmals um Faktor 20 aufzudrehen. Schließlich ist das eine absichtliche Störaktion, die ohne Nutzen unsere Ressourcen (oder vielmehr die der Foundation) verschwendet und sich nur noch quantitativ von einem DoS-Angriff unterscheidet.

Von der Hartnäckigkeit und langen Dauer der Aktion zeugen auch die beiden FzW-Einträge aus 2008, bei denen sich über die Statistik des Artikels gewundert wird: 1 und 2.

Nun zu meiner Frage: Was kann man tun? Gibt es vielleicht einen abuse Ansprechpartner der Foundation? Oder sollte man das einfach ignorieren?

Gruß, --Jodoform 18:11, 2. Feb. 2011 (CET)

> Gegenangriff ;-) --Nothere 22:12, 2. Feb. 2011 (CET)

Discover | Visualize | Dashboard | Settings | September 18th 2007, 19:53:23.433 to October 31st 2007, 07:49:50.025

(page.raw:myanmar OR page.raw:Than_Shwe OR page.raw:junta OR page.raw:Bhikkhu OR page.raw:"Aung_San_S

metrics

**Y-Axis**

**Aggregation**

Count

Advanced

buckets

**Split Lines**

**Aggregation**

Terms

**Field**

page.raw

**Order** **Size**

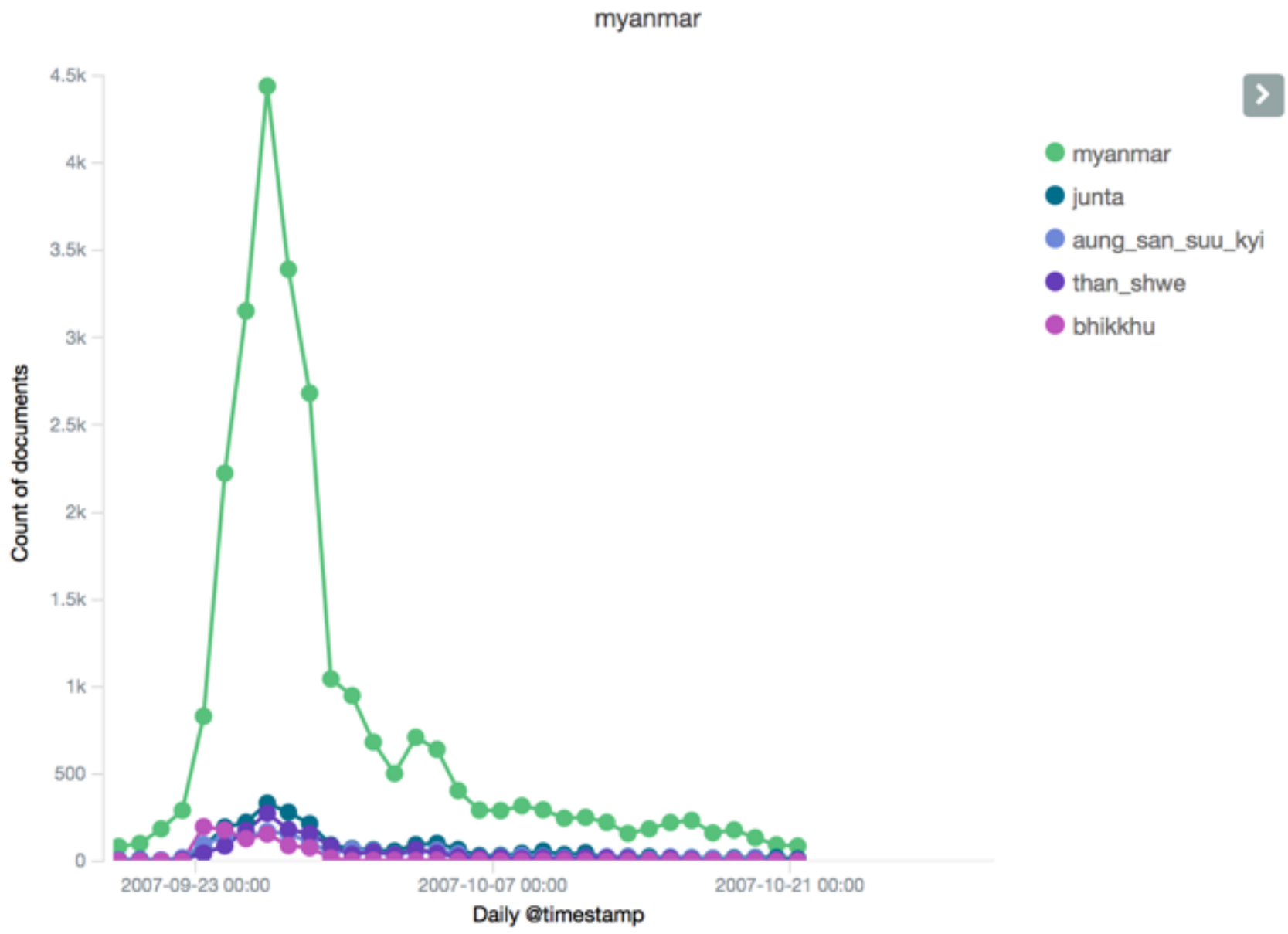
Top 51

Advanced

**X-Axis**

Apply

Discard



Discover | Visualize | Dashboard | Settings | September 18th 2007, 19:53:23.433 to October 31st 2007, 07:49:50.025

*

metrics

**Metric**

**Aggregation**

Count

Advanced

+ Add Aggregation

buckets

**Split Table**

Rows | Columns

**Aggregation**

Date Histogram

**Field**

@timestamp

**Interval**

Daily

Apply

Discard

terms per day wikipedia

**Daily @timestamp: September 18th 2007, 02:00:00.000**

| Top 5 page.raw | Count of documents |
|---|---|
| hauptseite | 10444 |
| spezial:zufällige_seite | 1812 |
| spezial:beobachtungsliste | 649 |
| jim_morrison | 491 |
| the_doors | 294 |

Export

Page Size All

**Daily @timestamp: September 19th 2007, 02:00:00.000**

| Top 5 page.raw | Count of documents |
|---|---|
| hauptseite | 122799 |
| spezial:zufällige_seite | 17382 |

Discover | Visualize | Dashboard | Settings | September 18th 2007, 19:53:23.433 to October 31st 2007, 07:49:50.025

*

**Field**

@timestamp

**Interval**

Daily

Advanced

**Split Rows**

**Sub Aggregation**

Significant Terms

**Field**

page.raw

**Size**

5

Advanced

Add Sub Aggregation

visualization options

Apply

Discard

terms per day wikipedia

**Daily @timestamp: September 18th 2007, 02:00:00.000**

| Top 5 unusual terms in page.raw | Count of documents |
|---|---|
| amyloidose | 284 |
| jim_morrison | 491 |
| the_doors | 294 |
| the_doors_(film) | 136 |
| howard_carpendale | 84 |

Export

Page Size All

**Daily @timestamp: September 19th 2007, 02:00:00.000**

| Top 5 unusual terms in page.raw | Count of documents |
|---|---|
| 19._september | 1213 |
| faustrecht_der_prärie | 1246 |

Discover | Visualize | Dashboard | Settings | September 18th 2007, 19:53:23.433 to September 19th 2007, 07:49:50.025

page.raw:"jim_morrison"

metrics

**Y-Axis**

**Aggregation**

Count

Advanced

buckets

**X-Axis**

**Aggregation**

Date Histogram

**Field**

@timestamp

**Interval**

Minute

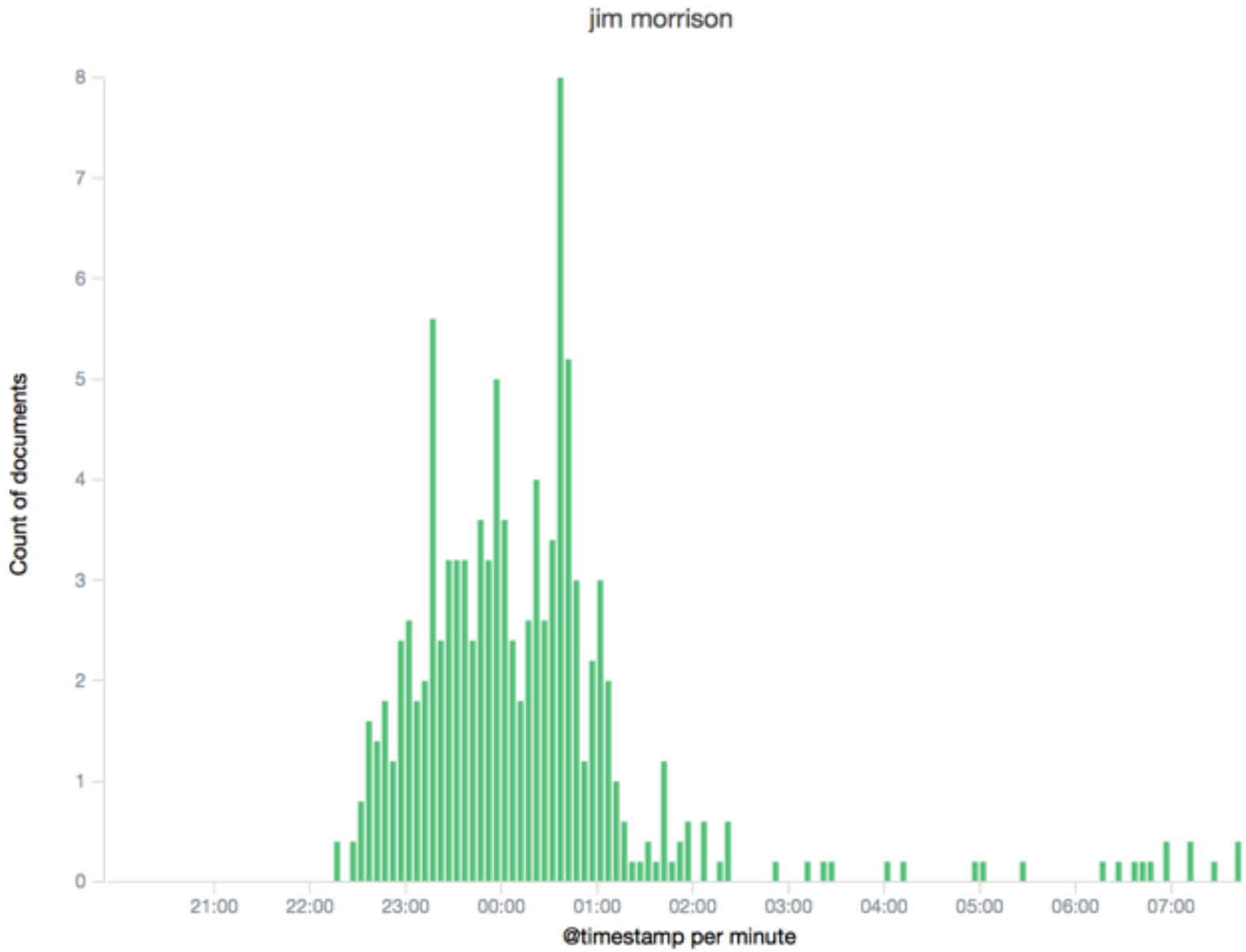
Advanced

Add Sub Aggregation

visualization options

Apply

Discard



⊞ Der Student Henry ist allein am Flughafen, um eine große Reise als Rucksacktourist anzutreten. In seiner Schüchternheit bringt ihn die kurze Begegnung mit einer schönen Reisenden völlig aus dem Konzept. Da bittet ihn eine ältere Dame, kurz auf ihren Koffer ...
(ARD/SWR)

**22:00** 16:9

ZIB 2

(ORF)

Seitenanfang

**22:25** 16:9

**Leben am Limit**

## The Doors

**Spielfilm, USA 1990**

**Darsteller:**

| | |
|---|---|
| Jim Morrison | Val Kilmer |
| Pamela Courson | Meg Ryan |
| Robby Krieger | Frank Whaley |
| John Densmore | Kevin Dillon |
| Ray Manzarek | Kyle MacLachlan |
| Cat | Billy Idol |
| Tom Baker | Michael Madsen |
| | u.a. |

Länge: 135 Minuten
Regie: Oliver Stone

⊞ Oliver Stone zelebriert die Ereignisse um die legendäre Rockband "The Doors" und ihres Leadsängers Jim Morrison als rauschhaftes visuelles und akustisches Filmereignis. Val Kilmer verkörpert die Rolle des in die Rockgeschichte eingegangenen charismatischen Sängers mit ...

**0:40**
VPS 21:50

**Nachrichtenmagazin des Schweizer Fernsehens:**

## 10 vor 10

# Terms aggregations can be expensive, so…

Check out:

- Fielddata filter + segment merge (optimize)
- Query cache
- Warmer
- Doc Values

Before moving to production, watch this video:

http://www.elasticsearch.org/webinars/elasticsearch-pre-flight-checklist/

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# Other interesting aggregations

What else?

- Percentiles
- Cardinality
- Significant terms

Coming soon:

- Derivatives?
- Correlations?
- …

https://github.com/elasticsearch/elasticsearch/issues/8110

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

# Thanks for listening!

# Q & A

## We're hiring!

http://elasticsearch.com/about/jobs

## We're helping!

http://elasticsearch.com/support

http://elasticsearch.com/training

Copyright Elasticsearch 2014. Copying, publishing and/or distributing without written permission is strictly prohibited

elasticsearch.